Volume 194, number 2 FEBS 3211 January 1986

A DNA binding protein from the extreme thermophile *Thermus thermophilus*

Rainer Zierer, Mathias Grote, Jan Dijk and Keith Wilson*

Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, Ihnestrasse 63–73, D-1000 Berlin 33, Germany

Received 25 October 1985

A DNA binding protein was extracted and purified from *Thermus thermophilus* HB8 grown at 75°C. The protein has an M_r of 10000 and from the partial data available shows a weak but clear sequence homology to the DNA binding protein II family. The protein is able to bind to DNA, and can also bind to RNA and to ribosomes. It affects the melting profile of the nucleic acids. From circular dichroic spectra an α-helix content of 32% and β-sheet of 40% are estimated. The proton NMR spectrum indicates a well folded tertiary structure. Upon binding of oligonucleotides pronounced changes in the NMR spectrum of the protein can be seen. Comparable changes are observed in circular dichroic spectra of protein-oligonucleotide complexes. Hexagonal crystals, which diffract to at least 3 Å, of the DNA binding protein have been obtained using ammonium sulphate as precipitant. The space group is $P6_222$, with cell dimensions $a=b=62.9$ Å and $c=110.6$ Å.

Thermophile *DNA binding protein* *Crystal*

1. INTRODUCTION

DNA binding protein II [1], which binds non-sequence specifically to double-stranded DNA, is ubiquitous in the eubacterial kingdom. It appears to have a histone-like role in the organization of the prokaryotic genome, possessing the ability to produce 'bead-like' structures which can be visualized in the electron microscope [2]. The amino acid sequence of the protein from a number of organisms (*Escherichia coli* [3] where there are 2 forms of the protein, *Bacillus stearothermophilus* [4], *Clostridium pasteurianum* [5], *Rhizobium meliloti* [6] and *Pseudomonas aeruginosa* [7]) has been determined. These proteins are highly homologous, with identical amino acids at more than 50% of the positions in the polypeptide chain, in spite of the divergence of some of the organisms from one another more than 3 billion years ago. A fuller discussion of the protein is given in [8].

* Present address: EMBL Hamburg c/o DESY, Notkestrasse 85, D-2000 Hamburg 52, FRG

The protein has also been isolated, sequenced and studied from an archebacterium: *Thermoplasma acidophilum* [9]. This protein has roughly one third of its residues in common with its eubacterial counterparts, including essentially all of those which are highly conserved within the latter.

The crystal structure of the DNA binding protein II from the moderate thermophile *B. stearothermophilus* has been solved at 3 Å resolution in our institute [8]. The dimeric structure is held together by a set of α-helices in close contact and a hydrophobic core composed primarily of phenylalanine residues. The helices support a surface of two 3-stranded β-sheets, one from each monomer. From these sheets protrude 2 extended arms of 20 amino acids. The first half of each arm lies in the β-ribbon conformation in the crystal, as an extension of 2 adjacent strands in each of the sheets. There is no electron density for the second half of the arms in the crystal, and these residues are presumed to be disordered. Using computer graphics, a model has been proposed for the

Published by Elsevier Science Publishers B.V. (Biomedical Division)

binding of the protein to DNA, involving the interaction of the extended arms with the complementary grooves on either side of the DNA double helix.

In addition to the group of DNA binding proteins II with what appears to be the same role in the various organisms, namely non-specific binding to and organization of the DNA, other workers have recently isolated and sequenced 2 independent proteins (or their genes) which are homologous to this group. These are the initiation host factor [10,11] and the transcription factor 1 [12] proteins. These have about 40% identical residues to the *B. stearothermophilus* DNA binding protein II. The 2 proteins show a degree of specificity in their DNA binding, and suggest the existence of an extensive family of DNA binding proteins based on the dimeric structural motif of DNA binding protein II.

We report here the isolation, solution properties and crystallization, of a homologous protein from the extreme thermophile *Thermus thermophilus.* This eubacterium can grow at temperatures in the ranges of 70–80°C [13]. We began our studies of *T. thermophilus* as a potential alternative source of thermostable ribosomal and DNA binding proteins [14] at a time when our crystals for the proteins from *B. stearothermophilus* were not of sufficient quality or size to allow a full three-dimensional (3D) analysis. Since that time the crystal quality has markedly improved allowing analysis of the *B. stearothermophilus* DNA binding protein II [8] and 3 ribosomal proteins [15–17].

However the DNA binding protein isolated from *T. thermophilus* shows interesting differences from the type II proteins previously studied. For the reasons discussed below it is an excellent candidate for a full sequence and crystallographic analysis which will provide further insight into the details of DNA-protein interaction in this important family.

2. MATERIALS AND METHODS

Cells of *T. thermophilus* strain HB8 were grown to late log phase and the ribosomes were prepared as described in [14]. The DNA binding protein was eluted after the ribosome fraction during gel filtration on an S-300 Sephacryl column. The DNA binding protein was further purified by chromatography on CM-Sepharose CL-6B as in [18].

2D polyacrylamide gel electrophoresis was performed as in [19].

Circular dichroic spectra were recorded using a Cary 60 spectrometer with a 1 mm path length cell. The cell compartment was maintained at the required temperature by water circulated from an external bath. The spectra were calibrated with camphorsulphonic acid. Binding protein at a concentration of 0.2 mg/ml was dialysed against 350 mM KF in 5 mM potassium buffer, pH 7.0. Urea denaturation was performed at 20°C. Samples of the same protein concentration were mixed with 350 mM KF in 5 mM potassium phosphate buffer containing concentrations of urea up to 8 M. The decrease in the ellipticity at 220 nm was used to measure the extent of the denaturation.

NMR spectra were recorded on a Bruker WH-270 spectrometer (Department of Biophysics, Portsmouth Polytechnic, England). Protein samples at 3 mg/ml in a volume of 0.3 ml were dialysed against 0.1 M NaCl in 10 mM sodium phosphate buffer, pH 7.0. Denaturation of the protein was carried out in the same buffer containing up to 8 M [^{2}H]urea. A mixture of dA_8/T_8 with protein in a molar ratio of 2:1 was used for the oligonucleotide binding spectra.

DNA binding properties of the protein were studied by filter binding assay [20] and gel filtration of nucleic acid (DNA, RNA or ribosomes) in 350 mM NaCl with 20 mM sodium phosphate buffer at pH 7.0 and also in polymix buffer [21]. The experimental conditions were as in [18].

The electron microscopy of nucleic acid-protein complexes was carried out as in [22] using plasmid pBR322 DNA. Binding protein and DNA in a weight ratio of 10:1 were incubated in 50 mM KCl, 0.5 mM DTE, 0.5 mM EDTA, pH 7.5 for 15 min at 37°C. Free protein was removed from the complex by Sepharose-4B column chromatography in the incubation buffer.

The protein was crystallized by the hanging-drop vapour diffusion method [23]. Precession photographs were recorded with CuK-α radiation, using an Enraf-Nonius camera on a Seifert stationary-anode generator.

3. RESULTS

The DNA binding protein was located in the

fractions emerging from the Sephacryl S-300 column after the ribosome peak by analysis on SDS gels and appeared as a prominent band of M_r 10000. Further purification was obtained by affinity chromatography on a heparin-Sepharose CL-6B column eluted with an NaCl gradient of 0.1–0.7 M. The protein was eluted from the column at 0.55 M NaCl (fig.1). At this stage it was judged pure as determined by SDS gel electrophoresis.

On 2D gels the protein showed a mobility distinctly different from that of DNA binding protein II of *E. coli* and *B. stearothermophilus* [18] as shown in fig.2. It appears in the region of small basic proteins and, in the pattern of 50 S ribosomal proteins, seems to comigrate with *T. thermophilus* protein L30. Whether protein L30 is in reality the DNA binding protein still bound to the ribosome cannot be decided at this stage.

The circular dichroic spectrum of the DNA binding protein is typical of a protein with a moderate amount of α-helix, as indicated by the value of the ellipticity at 220 nm (10000). A more accurate estimate of the distribution of secondary structure was obtained from an analysis of the spectrum using the programme CONTIN [24] which computes 32% α-helix and 40% β-sheet for the protein. The observed values for the known 3D structure of the *B. stearothermophilus* type II protein are 38% α-helix and 42% β-sheet, assuming the disordered arms in the crystal are indeed in a β-ribbon conformation.

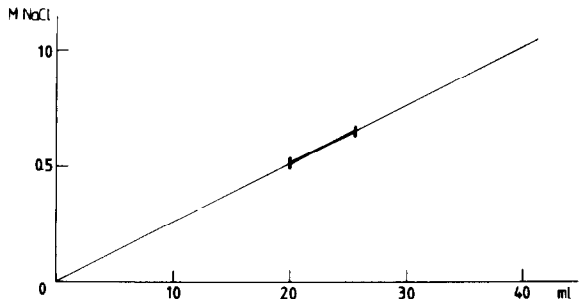

Fig.1. The position of the DNA binding protein in the eluent of the heparin-Sepharose Cl-6B column.

Fig.2. A 2D electrophoresis gel with the DNA binding protein superimposed on a background of *T. thermophilus* 50 S proteins.

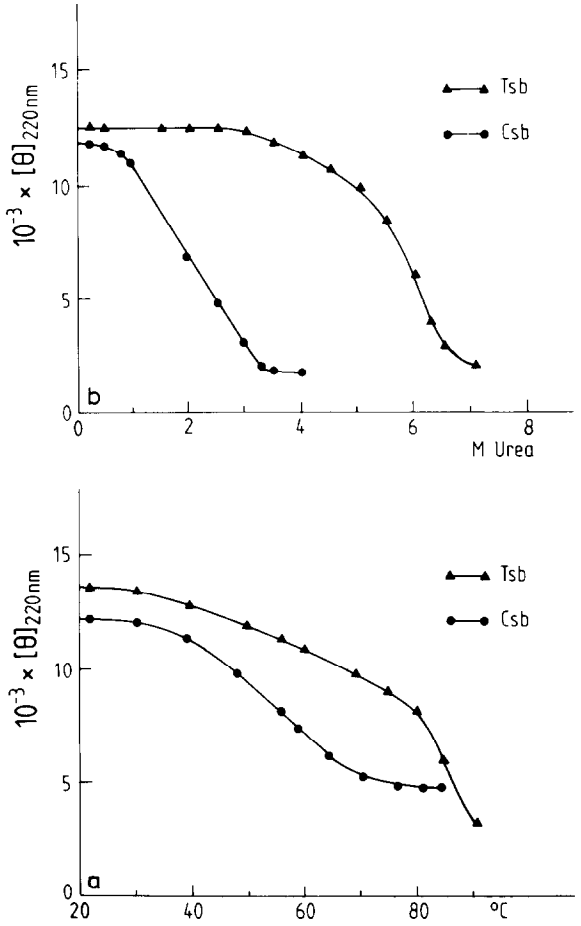

Fig.3. (a) The thermal denaturation of the DNA binding protein as followed by the ellipticity at 220 nm. (b) A comparable curve for the urea denaturation. The behaviour of the homologous protein (Csb) from the mesophile *C. pasteurianum* is also shown.

The resistance of the protein structure to thermal or urea denaturation was studied using the ellipticity at 220 nm as a monitor. As shown in fig.3 this value gradually decreases with increasing temperature up to 70°C followed by a sharp transition around 85°C. The midpoint of the urea denaturation curve occurs at 5.8 M (fig.3). In both cases the unfolding can be reversed by returning the protein to its original conditions.

The NMR spectrum also indicated that the protein has a well defined tertiary structure (fig.4). The spectrum of the native protein and that in 8 M urea are very different. Characteristic in the spectrum of the denatured protein are the sharp resonances of the CH_2 protons of lysine at 3 ppm and of the aliphatic methyl protons at 1 ppm. Three aspects of the native spectrum clearly indicate a well ordered fold: the non-exchangeable amide proton resonances between 8–10 ppm, the perturbed aromatic resonances between 6–8 ppm, and the poorly resolved ring current shifted methyl peaks between 0.3–0.7 ppm. The NMR spectrum recorded after the addition of a 2-fold molar excess of a mixture of dA_8/T_8 (fig.4), suggests a significant conformation change on nucleotide binding. Preliminary results from circular dichroic spectroscopy on such complexes support this observation. From the difference spectra (not shown) it can be concluded that conformational changes occur both in the nucleic acid component as well as in the protein.

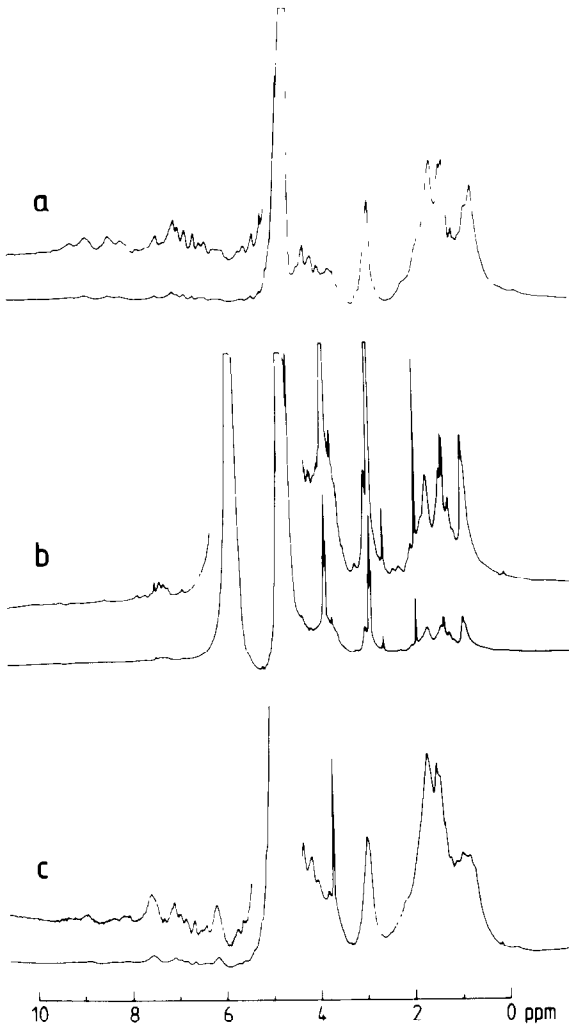

Fig.4. ^{1}H-NMR spectra of (a) native, (b) urea-denatured DNA binding protein and (c) the spectrum after addition of a dA_8/T_8 mixture.

In electron microscopy experiments on mixtures of the protein with plasmid pBR322 circular DNA, garland-like looped structures were observed (fig.5).

The binding to DNA was confirmed by gel filtration experiments in which the protein comigrated with the DNA. However in comparable experiments binding to ribosomes, ribosomal subunits, 16 S and 5 S rRNA was also observed. The effect of protein binding on the DNA structure was followed by the melting of the latter. DNA starts to melt at 60°C and shows a relatively sharp transition at 81°C. In the protein-nucleic acid complex the melting again starts at 60°C, but now a sharp transition is observed at 68°C, followed by a

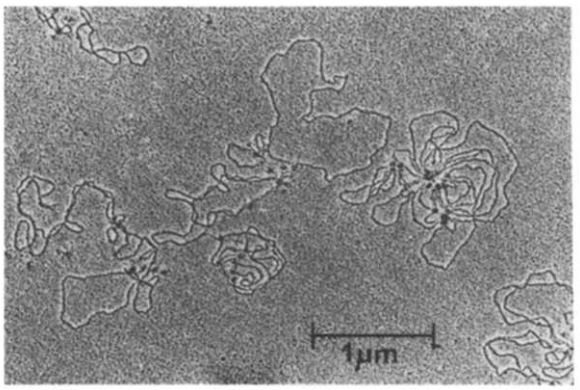

Fig.5. An electron micrograph of the complexes of the DNA binding protein with pBR322 plasmid DNA.

second, highly cooperative, melting step at 96°C (fig.6).

The protein was crystallized from ammonium sulphate using the hanging-drop vapour diffusion method. Over a period of 3 months at 20°C large bipyramidal crystals were obtained (fig.7) at an ammonium sulphate concentration of 3.95 M. The X-ray precession photographs indicate a hexagonal symmetry for the crystal with useful diffraction to at least 3 Å. The space group is $P6_222$ with cell dimensions $a = b = 62.9$ Å and $c = 110.6$ Å.

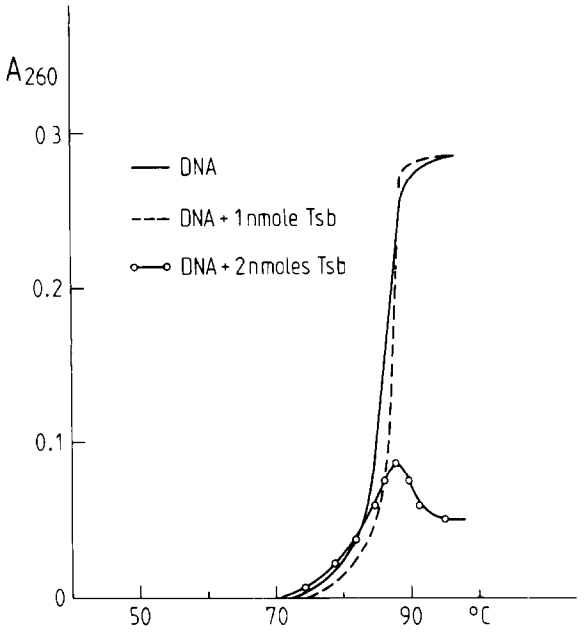

Fig.6. The melting curves for the DNA complexed with the protein.

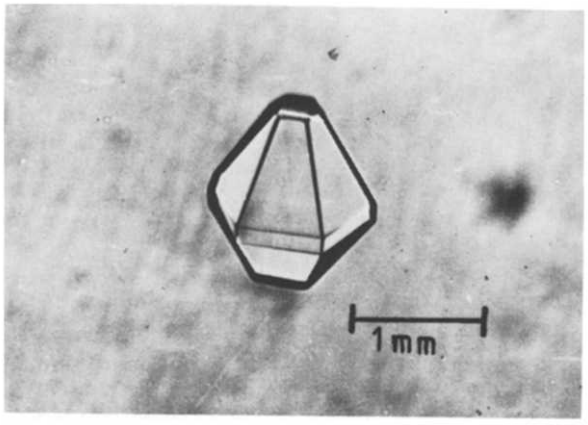

Fig.7. Hexagonal crystals of the DNA binding protein. The bar indicates a length of 0.1 mm.

4. DISCUSSION

DNA binding protein II has been isolated previously from a number of eubacteria and from one archebacterium. The common characteristics are that the protein exists as a dimer (monomeric M_r 10 000), is present in large amounts in the cell, carries a weak positive charge, and is both acid-soluble and heat-stable. The proteins isolated according to those criteria are related by sequence with strong conservation of residues 46–50 (-Gly-Phe-Gly-X-Phe-) and 61–65 (-Arg-Asn-Pro-X-Thr-) and have several other totally conserved positions scattered through the sequence [4].

We have isolated a DNA binding protein from *T. thermophilus* with the same M_r as the type II proteins using a similar extraction procedure. From the partial sequence data currently available (Zierer and Kimura, unpublished) the protein has a significant homology to the type II proteins, although several distinct differences are evident between this *T. thermophilus* proteins and the type II proteins of other bacteria: (i) the protein is more basic than the other type II proteins; (ii) Several of the individual totally conserved residues in the N-terminal half of the type II proteins are not conserved in the *T. thermophilus* protein, whereas there is a strong homology in the C-terminal half; (iii) the NMR spectrum of the protein is significantly different from that of the *E. coli* and *B. stearothermophilus* type II proteins [18,25] suggesting there may be some differences in the 3D structure. These spectral changes may, of course, merely reflect the considerable amino acid substitutions accommodated in a similar tertiary fold; (iv) as observed by NMR spectroscopy the structural changes accompanying oligonucleotide binding are very small for the *E. coli* and *B. stearothermophilus* type II proteins, and are restricted to an upfield shift of the arginine resonances [18,25]. This is in keeping with the mode of binding inferred from the 3D structure which involves the positively charged arginine residues present in the 2 flexible arms wrapped around the DNA [8]. The changes reported here for the spectrum of the *T. thermophilus* protein are much

larger, perhaps indicating a larger conformational change of the protein on binding.

Given these differences it is not absolutely clear whether the *T. thermophilus* protein is indeed the type II protein of that organism. The determination of the complete amino acid sequence is in hand, and this will allow the establishment of a more precise relationship between this protein and the type II group, and indeed the initiation host factor and transcription factor 1 proteins. Further searches for the presence of a more typical type II protein in the organism will be carried out.

A more detailed understanding of the structure and function of the protein and the mechanism of its interaction with DNA must await determination of the complete sequence and, ultimately, of the 3D structure. The latter is of particular interest, partly because the protein appears to be somewhat different from the classic members of the type II group.

The crystal form of the *T. thermophilus* protein confers considerable extra interest on the 3D analysis. In the *B. stearothermophilus* protein crystals one half of the extended arms of the protein is disordered and also the proposed DNA binding site is inaccessible to oligonucleotide binding as it is occupied by the adjacent protein molecule along the *b* axis. This occurs for all 3 independent molecules in the cell. The conformation of the arms and their proposed interaction with the DNA grooves have thus relied on computer model building rather than experiment. Moreover we have several other crystal forms of the *B. stearothermophilus* protein but all possess a cell dimension of about 37 Å, the *b* cell dimension of the crystals already analysed, suggesting that this method of packing may be common to all the crystals (Appelt, Tanaka and Wilson, unpublished).

The cell dimensions of the *T. thermophilus* protein crystals give a packing volume, V_m, of 1.66 $Å^3$/Da for a dimer, and of 3.32 for a monomer in the asymmetric unit. The latter is almost certainly the true value and corresponds to a rather loose packing of the molecules in the lattice, the value being at the upper end of the range normally observed in protein crystals [26]: the V_m for the *B. stearothermophilus* crystals is 2.7. The expected extra thermostability of the *T. thermophilus* protein and the different, and looser, packing of the molecules may give access to the binding site and the arms of the protein may have ordered conformation in these crystals.

ACKNOWLEDGEMENTS

We should like to thank Drs S.W. White, K. Appelt and W. Bennett for help and discussion, and Professor H.G. Wittmann for critical reading of the manuscript.

REFERENCES

[1] Geider, K. and Hoffmann-Berling, H. (1981) Annu. Rev. Biochem. 50, 233–260.
[2] Rouviere-Yaniv, J., Yaniv, M. and Germond, J.E. (1979) Cell 17, 255–274.
[3] Mende, L., Timm, B. and Subramanian, A.R. (1978) FEBS Lett. 96, 395–398.
[4] Kimura, M. and Wilson, K.S. (1983) J. Biol. Chem. 258, 4007–4011.
[5] Kimura, M., Kimura, J. and Zierer, R. (1984) FEBS Lett. 175, 208–212.
[6] Laine, B., Belaiche, D., Khanaka, H. and Sautiere, P. (1983) Eur. J. Biochem. 131, 325–331.
[7] Hawkins, A.R. and Wootton, J.C. (1981) FEBS Lett. 130, 275–278.
[8] Tanaka, I., Appelt, K., Dijk, J., White, S.W. and Wilson, K.S. (1984) Nature 310, 376–381.
[9] Delange, R.J., Williams, L.C. and Searcy, D.G. (1981) J. Biol. Chem. 256, 905–911.
[10] Flamm, E. and Weisberg, R. (1985) J. Mol. Biol. 183, 117–128.
[11] Miller, H.I. (1984) Cold Spring Harbor Symp. Quant. Biol. 49, 691–698.
[12] Green, J.R., Brennen, S.M., Andrew, D.J., Thompson, C.C., Richards, S.H., Heinrickson, R.L. and Geiduschek, E.P. (1984) Proc. Natl. Acad. Sci. USA 81, 7031–7035.
[13] Oshima, T. and Imahori, K. (1974) Int. J. Syst. Bacteriol. 24, 104–112.
[14] Zierer, R., Wilson, K.S. and Dijk, J. (1985) in preparation.
[15] Appelt, K., Dijk, J., White, S.W., Wilson, K.S. and Bartels, K. (1983) FEBS Lett. 160, 72–74.
[16] Appelt, K., Tanaka, I., White, S.W. and Wilson, K.S. (1984) FEBS Lett. 165, 43–45.
[17] White, S.W., Appelt, K., Dijk, J. and Wilson, K.S. (1983) FEBS Lett. 163, 73–75.
[18] Dijk, J., White, S.W., Wilson, K.S. and Appelt, K. (1983) J. Biol. Chem. 258, 4003–4006.
[19] Geyl, D., Böck, A. and Isono, K. (1981) Mol. Gen. Genet. 181, 309–312.

[20] Lammi, M., Paci, M. and Gualerzi, C. (1984) FEBS Lett. 170, 99-104.
[21] Jelenc, P.C. (1980) Anal. Biochem. 105, 369-374.
[22] Pueler, A. and Timmis, K.N. (1980) Advanced Molecular Genetics, pp. 281-302, Springer, Berlin.
[23] Davies, D. and Segal, D.M. (1971) Methods Enzymol. 22, 266-269.
[24] Provencher, S.W. and Glöckner, J. (1981) Biochem. 20, 33-38.
[25] Paci, M., Pon, C.L., Losso, M.A. and Gualerzi, C. (1984) Eur. J. Biochem. 138, 193-200.
[26] Matthews, B.W. (1968) J. Mol. Biol. 33, 491-497.